

**Summary Report**  
**Northeast Fisheries Science Center Science Data Collection Program Review**

Reviewer 1

On August 5-9, 2013, a review was conducted of the stock assessment data collection programs of the Northeast Fisheries Science Center (NEFSC). Six panelists reviewed the program following the terms of reference provided in Appendix I. Feedback was based on presentations during the visit and background documents provided in advance of the review ([http://www.nefsc.noaa.gov/program\\_review/agenda.html](http://www.nefsc.noaa.gov/program_review/agenda.html)). The review focused on fishery-independent and fishery-dependent data, data management, and quality control as they relate to stock assessments for which the NEFSC has jurisdiction. Presentations provided an overview of each topic, as well as perceived strengths, weaknesses, and potential solutions. Each panelist produced an independent evaluation including suggestions for improvement and new initiatives. As Chair of the review panel, I was asked to read each panelist's report and summarize common themes.

We thank the NEFSC staff for their work in preparing materials for this review and for the helpful discussions during our visit. Our goal is to draw on our own experiences and the material provided, and give constructive feedback that will lead to improvements in science and management. Provided below are some of the major themes, but individual reports provide much more detail regarding potential refinements of current methods and ideas for new approaches:

- The NEFSC has a world-class survey program that supports stock assessment for more than 45 species. The long time series is a significant asset to the Center's stock assessment scientists (and for wider regional applications in marine ecosystem understanding). A sophisticated data management system reduces the risk of error and increases the timeliness of providing survey data to assessment scientists.
- Panelists made several recommendations for improving the peer review process. The most common suggestion was to ensure that presentations and background documents were at a level consistent with the terms of reference and the scope of the review.
- Trip-specific identifier. Presenters made clear the difficulties of linking trip and catch information, which is critical for the sophisticated management system that the data streams are intended to support.
- Management strategy evaluation (MSE). This would allow for a variety of changes to be explored, such as a reduction in trawl survey frequency or changes in sampling intensity for ages. The CV achieved for a particular survey or monitoring program (e.g. CV for age composition or discard) is not an endpoint; what matters is how precision affects the information used to manage each stock. An MSE could be used to explore current complex versus simpler assessment and management approaches.
- Allocation of NEFSC resources. Various schemes could be used to allocate time and funds (e.g. based on the value of the resource, precision of assessment results, risk of stock collapse). A formal analysis with clear criteria should be useful in defending funding levels and in planning future budgets and staffing.
- Allocation of NEFSC staff time between routine work and research that has potential for substantial gains in efficiency or data reliability. For example, age validation work takes time away from production aging but could identify significant biases that negate the value of production aging. Another example would be to work on estimating survey catchability, so that surveys provide absolute rather than relative abundance information. Staff also need time for strategic thinking and publishing. Providing time for professional development can aid in retaining key staff who might otherwise be lost due to burnout.

- Allocating more NEFSC resources into new methods that have potential to substantially improve assessment precision and accuracy. HabCam was mentioned by most or all panelists as a highly innovative tool that is providing highly valuable data for scallops and may pay off for multiple additional species. The key is to provide the time and resources to develop these or other new methods (e.g video, AUVs, new acoustic methods). Acoustic surveys have potential to enhance interpretation of existing bottom-trawl datasets, and new surveys such as the integrated pelagic survey can improve single-species assessments and lead the way toward ecosystem-level models and management. The NEFSC has the capability (in-house and nearby) to make important advances through acoustic surveys but the capability seems underutilized. Similarly, the NEFSC has tremendous resources (data and staff) in food web modeling that (given time for assessment scientists to pursue) could make a much greater contribution to stock assessments.
- Developing a probability-based sampling scheme for fishery-dependent monitoring, particularly port sampling. The MRIP program was put forth as an example where sufficient time and expertise allowed for a completely revamped survey.
- Better integration of NEFSC and state surveys. This could include planning efforts to standardize timing and methods, to improve comparability among surveys. On the stock assessment side, panelists questioned the appropriateness of giving equal weight to a survey covering the whole range, compared to a large set of geographically restricted surveys of unknown rigor.
- Well designed observer program, ideally should be expanded in coverage. While obviously expensive, this has potential to provide better data of the spatial scale that is desired by management.
- Cooperative research surveys need to be carefully designed, so that the data will be used in assessments.
- Pelagic resources should receive greater emphasis, not only to support stock assessment and management but also to facilitate the shift toward ecosystem management. The new integrated pelagic survey is an important step and might have been a useful topic for inclusion in this review.

## Northeast Fisheries Science Center Science Data Collection Program Review

Reviewer 1

Comments on the review process:

- The review is a substantial time commitment when the week on-site is combined with the necessary time beforehand to review documents. Potential panelists should be advised of the full scope of the commitment up front.
- Documents provided to the panel should be at an appropriate level for addressing the terms of reference. Documents described as core were generally quite detailed, covering the details of carrying out the task (survey or monitoring program) rather than addressing the program's impact and scope. For future reviews, it might be better to have staff focus on the terms of reference and produce a summary document that better described the scope of the program and how its data contribute to stock assessment accuracy and precision.

Fishery independent:

The NEFSC has a long-standing survey program that provides very high quality data. Survey data (e.g. CPUE, length composition and age composition) from the spring and fall bottom trawl surveys are used in more than 45 stock assessments. This is a clear indication of their importance, but still does not answer the TOR question of what each survey contributes to assessment precision and accuracy. One crude way of assessing each survey's worth would be to examine the typical sensitivity analysis for an assessment, with the survey included versus excluded. Seeing how the assessment precision changed with and without each NEFSC survey would begin to address this question. This approach could also be used to determine the relative contribution of the spring versus fall surveys, regarding the need for two surveys per year. This question was posed during the panel review, and the response was that the spring survey provided information about the adult stock whereas the fall survey had better information on recruitment. However, catch-at-age data are often very informative about recruitment so that may not be a strong reason to retain the fall survey. Accuracy (TOR#1) is very difficult to assess because the right answer is unknown, but converged past estimates might be useful as a "relatively well estimated" benchmark for examining estimates with and without each survey. Survey results including age data are available for assessments in a very timely manner (TOR #1). Some of the data management procedures (e.g. barcodes for otolith tracking) are innovative and promote quality data and timeliness.

One area that might be given more weight moving forward is more basic research on new methods. The HabCam is an example of how a new survey method can improve assessment precision and accuracy. It provides absolute abundance estimates and can be used to survey large areas, compared to the traditional dredge which covers less area and underestimates true density. Allocating some resources to pilot studies of new methods (e.g. video, an expanded role for acoustic surveys) would seem worthwhile, in addition to maintaining traditional data streams. One potential advantage of new methods such as video or acoustics is that it is sometimes possible to measure absolute abundance. Compared to indices of relative abundance, absolute abundance data provide a much stronger constraint on a stock assessment and could help in resolving natural mortality (which is often simply based on an assumed value). Using an assumed value (life history correlate) makes the assessment results conditional on that assumption. If the assumed  $M$  is wrong or changes through time because of temporal factors such as predator abundance, then precision is irrelevant because of the unknown bias.

A substantial effort has gone into survey standardization. Among the important innovations are the net mensuration data used to characterize net geometry and to reject/repeat invalid tows. An intensive and peer-reviewed calibration study was done with the transition to a new research vessel and new net design.

One possible direction for improving the value of the survey data would be through spatially explicit (geostatistical) models. Habitat data from multibeam sonar could be coupled with the many years of trawl survey data to develop models that take into account depth, topography, bottom type and other factors that affect fish abundance. Spatially explicit models might provide an avenue for using fishery knowledge (and perhaps the industry-selected tows in cooperative surveys) and should improve the precision of survey estimates. Models that capture more detail about fish distribution might improve industry acceptance of survey methods, and could be an avenue for future stratification schemes that take into account fish distribution patterns.

It is not clear how the NEFSC's resources (staff time, funding) are allocated among competing needs, and a formal scheme for allocating resources could be useful. For example, how much staff time should be devoted to traditional stock assessment versus developing new models (e.g. an assessment that explicitly incorporate predation mortality) or production aging versus age validation? Staff time for production aging is scheduled to meet assessment needs, but it appears to be difficult to carve out time to do age validation work (which could show that production aging was producing reliable or biased results). Questions about allocating NEFSC resources also came up regarding the scallop survey. The scallop fishery has very high value and is reported to be thriving due to quality survey data and effective management, but it has apparently been difficult to get survey funding. This is an innovative survey incorporating a new gear (HabCam) that has provided excellent abundance and size data and reduced the number of survey dredge tows. In addition to value, NEFSC resources could also be allocated based on stock status (allocating resources based on the size of the resource or perhaps stocks in decline) or assessment precision. It is hard to compare assessment precision among species/stocks, however, because assessment estimates are often conditional on fixed parameters (e.g. an assumed  $M$ ).

One example of the struggle between traditional methods versus more basic research is regarding the use of food habits and consumption data. The Councils obviously have a constant need for information, but investing some time in more basic research might result in models that are more biologically meaningful and perhaps providing better management advice. Jason Link described the exemplary food habits database and described some of the limited ways those data have been used in assessments. The consumption estimates provide a way of ground-truthing predator and prey abundance estimates as well as estimates of the magnitude of natural mortality. In at least one example, the consumption estimates showed that the traditional (constant  $M$ ) stock assessment was biased and resulted in a more realistic (assumed) pattern for  $M$ . It is important to carve out time for stock assessment scientists to test new assessment approaches and non-traditional data sources. Stock assessments that incorporate predation mortality (even as a check on assumed  $M$ s, but even better by incorporating predation as a loss matrix) are more mechanistic, can reduce bias and uncertainty, and are an important step toward ecosystem management.

Regarding limitations/weaknesses of fishery independent data, one area that was not discussed was survey catchability. Basic research on this topic has potential to improve assessment precision and accuracy, and could aid in the transition from assessments based on assumed  $M$  values to ones in which  $M$  is estimated internally. Hydroacoustic monitoring which has been done for more than ten years during the spring and fall surveys would seem to be a potentially valuable data source for exploring catchability (and the vertical distribution of fish and their vulnerability to the trawls). It seems likely that a combined acoustic/trawl analysis would be more informative about population abundance and trends than using only the trawl data.

#### Fishery dependent:

The NEFSC's fishery dependent monitoring program is highly complex, as a result of the highly complex Council management schemes that the program is intended to support. There is a general tendency towards more complex management over time, and in this case, the NEFSC has had to develop complex systems for providing spatially-explicit data to meet those management needs. Given the likelihood of shrinking budgets in the future, it will be critical for NEFSC scientists and administrators to provide feedback to managers (e.g. Councils) about what is sustainable and worthwhile. We were not presented any results demonstrating the benefits of area-specific management, but it may be important to start examining the losses (and perhaps

gains) of simpler management approaches. One of the documents provided to the panel described a working group that is looking at the potential gains and losses of using simpler models for assessment and management. Continued work on that topic is recommended, because the NEFSC seems to be close to maximum capacity for meeting the data (and assessment) needs of the Councils. If future budgets are reduced, simpler assessment and management approaches may be unavoidable, so it would be wise to explore those options now.

The monitoring systems that have been put in place are providing valuable detail about fishing practices. Whether done by the NEFSC or other entities, there seems to be steady progress towards electronic reporting of effort and catch information. It is important to make reporting as painless and error-free as possible given the complex regulatory schemes under which fishermen operate. The NEFSC has also implemented many error checking protocols to catch various data entry errors (e.g. invalid fish lengths, lat/lon coordinates inconsistent with reported statistical area).

The NEFSC has put a lot of emphasis on quantitative evaluation of their fishery-dependent monitoring programs (e.g. effectiveness of observer monitoring program; prioritization of assessment updates based on change from MRFSS to MRIP; validation of catch and effort allocation to statistical areas). The prioritization of updates re the MRFSS/MRIP transition was interesting in that relative ranks were used for the 16 New England species (MRFSS/MRIP Calibration Workshop Ad-hoc Working Group Report. 2012). A similar quantitative ranking scheme with various criteria (size of resource, value of resource, level of assessment precision) could be useful for allocating monitoring effort among the stocks/fisheries for which the NEFSC has jurisdiction. The NEFSC has also examined the level of monitoring required to achieve a specific level of precision in discard estimates. This sort of exercise is important in budget planning and in determining what types of monitoring are not affordable, either now or if budgets are reduced in the future.

The process of assigning landings and effort to statistical areas is complex and has potential for at least moderate error, although it is likely to be less subjective and error prone than the prior method of using port sampler interviews. The new process of assigning landings and effort by area includes estimation of error, although it was not clear whether the variance estimates include uncertainty due to area assignments when trip level matching was not possible.

#### Cooperative research surveys

The NEFSC has devoted considerable resources to industry-based or cooperative surveys. These have the potential to improve communication lines between scientists and industry, and provide a way for industry knowledge to feed into the survey and assessment processes. Past efforts have been a combination of surveys and gear evaluations. Short-term studies to evaluate gear modifications or surveys of areas/species not addressed by the standard NEFSC surveys appear to have the most potential. It may be challenging to use industry vessels for long-term surveys because industry vessels are often upgraded in terms of gear designs or electronics. Another struggle in cooperative surveys is the conflict between fishing and a statistically valid survey. However, this can be an opportunity to increase understanding about research methods.

#### Assessment examples

The chosen assessments illustrate the range from data poor to data rich assessments, but it was difficult to take four specific examples and judge the scope and significance of the NEFSC's role in providing data for assessment and management. At least two of the four seem to be special cases that provide limited insight into the breadth of NEFSC's mission. The tilefish resource is small (landings of 1,000 t, \$5M in value), highly restricted in geographic scope and not well covered by the NEFSC trawl surveys. The herring example is a much larger and more valuable fishery (101,000 t, \$26M in value) but is a pelagic stock that is not well sampled by bottom trawl surveys. It might have been better to have used as examples more of the traditional species that have been the primary focus of the NEFSC (e.g. cod, haddock stocks).

Summer flounder is an example of a success (TOR #2) for the NEFSC regarding fishery independent data contributing to a strong assessment. There are multiple state surveys that cover

portions of the range, but the nearly rangewide NEFSC surveys play a central role. NEFSC survey CPUE data show that the stock reached a low point in around 1989, and has rebuilt continuously since that time. This species is a good example of how  $F$  can be brought down to a target level (and population size increased) through steady long-term management supported by high-quality survey data.

The panel requested (and received) additional information that provided a clearer picture about the breadth of the NEFSC's role in supporting stock assessment and management. One reason for requesting more "big-picture" information is that it was not clear how the NEFSC's efforts were allocated among species/stocks/fisheries. The spreadsheet shows that NEFSC bottom trawl data are used in 49 stock assessments. Length data contribute to 60 assessments and age data in 29 assessments. Some of these are relatively small fisheries in terms of landings or value but may be important from the perspective of rebuilding depleted stocks.

## Summary and Conclusions

- Allocate some NEFSC resources toward new survey methods that have potential to make substantial improvement in assessment reliability (accuracy and precision)
- Explore spatially-explicit models of fish density, based on the spring and fall bottom trawl surveys and incorporating environmental data and sonar data on topography and bottom type
- Allocate more NEFSC resources to basic research that has potential for substantial increase in knowledge (e.g. validation studies to support/improve production aging, stock assessment models with explicit inclusion of predation mortality)
- Consider a formal scheme for allocating NEFSC resources (criteria could include fishery value, size of the resource, risk of triggering management action, precision of assessment results)
- Evaluate gains and losses of simpler management approaches that are less data-intensive.

## Northeast Fisheries Science Center Science Data Collection Program Review

Reviewer 2

The Northeast Fisheries Science Center convened a review as part of a national effort to review each regional science center annually. The focus of the review in 2013 was to evaluate fishery-dependent and fishery-independent data as it relates to fishery stock assessments. The terms of reference included:

1. To what extent do fishery independent survey data quality, statistical precision, and timeliness impact overall assessment accuracy, precision and timeliness?
2. What are the major fishery independent data survey successes and how should they be supported?
3. What are the major fishery independent survey limitations/weaknesses and how could they be resolved? Define potential improvements and priorities for recommended improvements.
4. To what extent do fishery dependent data quality, statistical precision, timeliness issues impact overall assessment accuracy, precision and timeliness?
5. What are the major fishery dependent successes and how should they be supported?
6. What are the major fishery dependent data limitations/weaknesses and how could they be resolved?
7. What recommendations do you have for prioritizing fishery independent and fishery dependent data collection improvements?
8. To what extent are fishery independent and fishery dependent data readily accessible to Center stock assessment scientists and to various external researchers who may wish to replicate NMFS stock assessments?
9. Identify the highest priority needs for improving fishery dependent and fishery independent data. Define potential improvements.

It is difficult to answer many of these TOR completely, but somewhat more tractable to address the overarching questions given:

- Relationship of current and planned fishery assessment data activities to Center fishery assessments mandates and requirements – is the Center doing the right things?
- Opportunities – are there opportunities that the Center should be pursuing in collecting and compiling fishery assessment data, including shared approaches with partners?
- Scientific/technical approach – are the Center's fishery data objectives adequate, and is the Center using the best suite of techniques and approaches to meet those objectives?
- Organization and priorities – is the Center's fishery data system properly organized to meet its mandates and is the allocation of resources among programs appropriate?
- Scientific conduct – are the Center's fishery data programs being conducted properly?

It was difficult to determine the amount of integration among the various divisions of NEFSC. While a lot of detail was provided on specific programs, there was little to no discussion of how the pieces fit together and mechanisms in place at NEFSC that fostered collaboration, integration, and common planning. A diagram was used that provided boxes with various components of the stock assessment process, with arrows linking the boxes. While much information was presented about the boxes in the diagram, this reviewer would have appreciated more discussion about the interactions between data providers and stock assessment scientists, as well as interactions between fishery-independent data surveys and NEFSC ecosystem surveys.

It is evident that the NEFSC is tasked with huge responsibilities, and they are making huge efforts to meet the demands of changing and increasing management requirements. In some instances, the best efforts of the talented people at NEFSC are just keeping the data collection and

assessments going, leaving little room for substantial improvement or strategic thinking about the best way to approach things over the long term. These constraints are understandable, and the recommendations that follow are given with the full understanding that some of them may not be practical to implement immediately, given demands on personnel and budgets.

#### Fishery independent data

The NEFSC has built an extremely impressive fishery independent data collection effort around the groundfish surveys. The long time series and statistical design allows for standardization and consistent analysis, which is a significant asset to the Center's stock assessment scientists (and for many other wider regional applications in marine ecosystem understanding). An enormous amount of effort goes into the conduct and data management of the groundfish survey, which may not appear to be completely warranted given the total value of the catch in relation to other species such as sea scallops. However, the "value" of the observations needs to be evaluated in terms that include more than the monetary value of the fishery (or fisheries) in which the data is used for stock assessments. While it is easy to maintain that the trawl survey is "invaluable", some specific and strategic thinking about real costs and benefits might be warranted. This would also be helpful in budget battles when expenses are scrutinized.

The panel was not provided budget details on what NEFSC was spending on different aspects of the data collection and stock assessment process. While it was not in our charge to evaluate the balance of investments across the different activities at the NEFSC, it would have been instructive to see just a pie chart of percentages of the total NEFSC resources that went towards various parts of the whole. For instance, the data management system was presented as increasing efficiency, but it was unclear what those saved resources went into.

The history and political pressures surrounding groundfish issues in the region will continue to make the NEFSC trawl surveys a key data resource. Nevertheless, the NEFSC should continually evaluate appropriate survey methodology. Is it completely necessary to have 2 seasonal tows? What benefits are provided by the 2 surveys that one survey cannot satisfy? It may be possible to reallocate stations without losing value of the observations. Furthermore, sampling the same locations over the length of the time series may not really be sampling the same biological communities. Especially as temperatures warm and phenologies change, the survey sites may be either in the wrong place or at the wrong time to make valid comparisons with past years. This is a difficult problem to deal with, but the survey data should always be evaluated with these types of changes in mind. Interaction with the NEFSC ecosystem processes division could be useful in this regard. An interesting exercise might be to look retrospectively at past regime changes (like North Atlantic Oscillation shifts) and how stock assessment accuracy might have changed over the change in environmental conditions.

It should be made explicit that the trawl survey is not able to sample particular species effectively, and other approaches may be needed to sample the full array of species. Acoustic methods may be useful in this regard, as was presented to the panel on the final day of input. Overall, the NEFSC is not capitalizing on the investment that NOAA made in an acoustic survey capability on the Bigelow. Other panelists are more qualified to provide specific comments on the acoustic surveys. Optical methods such as HABCAM are useful for some species, but not all. The utility of these alternate sampling methods as well as other new technologies should continue to be evaluated. In addition, the monitoring needs of today may not be the needs of tomorrow. This has been seen recently as changing management regimes have demanded more and different information, and the NEFSC has struggled to keep up. As FMCs move towards more ecosystem approaches, a wider array of species may need to be sampled, and the trawl survey will have to be supplemented with other surveys. The integrated pelagic survey that was mentioned on the final day of input is a good step towards finding out what other surveys may eventually be required to



form a comprehensive idea of ecosystem processes. Sampling of pelagics is an obvious hole in the current sampling that was presented to the panel.

The recent switch to a new survey vessel and net system was carefully evaluated and calibrated. One drawback to the Bigelow is the inability to survey shallower waters. There is a need for a dedicated shallower-draft vessel that could successfully sample the inner stations across the entire sample domain. This would be preferable to the current approach, which is to use state-supported sampling that is not sampling on the same scales, or the same gear. It is difficult to determine what value the state surveys are in the stock assessments, since they survey only a small domain at different times. They may be appropriate for some assessments, but the panel was not provided overall information about how much how much particular species assessments changed with the addition/elimination of the state survey data. Perhaps this is in the individual stock assessment documents, but an overall evaluation could be useful.

Regarding managing the data collected on the trawl surveys, the NEFSC is to be commended for their truly stellar approach to tracking samples from collection through processing to incorporation into the database. On-the-fly tow evaluation and QA/QC is extremely impressive. It was evident from other presentations that the data management has been a valuable improvement. These efforts should continue to be supported, and NMFS should consider expanding the system to other FSCs, and explore possibilities for using the same type of rigorous system for dockside biological sampling.

It was difficult to judge how many samples were enough for proper aging. The aging process is lengthy and it was astounding to this reviewer that only one person aged a particular species. This seems to be an extremely precarious situation for a vital link in the assessment process.

The 2009 ACL working group white paper provided to the review team stated: “*Relative sources of uncertainty in the data and the cost effectiveness of acquiring data for assessments and Assessment Evaluations need to be evaluated. This could include such activities as statistical analyses of optimal sample sizes (N) of otoliths and scales required to support assessments*”. This recommendation still stands as an extremely important aspect of prioritizing sampling. The panel was not provided an update to this 2009 recommendation with regard to actions taken between the 2009 report and the present review.

#### Fishery dependent data:

The fishery dependent data collected by the NEFSC is complex and challenging to deal with. Much of the utility of the FDD is constrained by the challenges in linking VTR, VMS, and dealer data. This came up time and again in the presentations. Tremendous efforts are undertaken to come up with work-arounds that try to deal with this problem. But, as one presenter noted, these may be peer-reviewed and robust, but they are not solutions. The NEFSC absolutely needs to provide a solution to this instead of using large amounts of staff time and expertise to work around it. The news that a FDD working group is being set up is a good step towards solving this problem, and the working group should focus on this issue without digression into other topics until a solution is proposed. Some of the issues are no doubt historical, and the fact that responsibilities are shared between NERO and NEFSC doesn't help matters. Integration of effort between the NERO and NEFSC will have challenges, but it is unacceptable that this situation cannot be resolved.

The port biological sampling program does not seem to have the same degree of a statistical design as the recreational fisheries sampling. While the recreational fisheries data collection program has been reviewed by the National Academy of Sciences (and apparently will be again in the near future), it was not clear if other FDD sources had the benefit of similar reviews. This

reviewer does not have the appropriate expertise to provide a useful evaluation of the port biological sampling design, and other reviewers may be able to provide more input on this issue. The same comment that was made above re. optimizing the numbers of samples collected for otolith analysis applies to the port biological sampling.

The observer program provides important data that cannot be collected through any other means. The training and standards for observers is of high quality, and the data audits and editing provide QA/QC. The direct link with industry was identified as a key strength. While the filtering process for sea day allocation makes sense given constraints in funding, better coverage of observers is needed. Support should be sought to increase the number of sea days. It may be possible to look to industry to cost-share in this regard, especially if fishermen can see the benefits under new management structures.

Industry collaboration in data collection is an important way to engage stakeholders and foster better communication and mutual understanding. Some progress has been made in this regard. However, the utility of industry-collected data in the stock assessment process seems fairly low. Great care needs to be taken to provide realistic expectations to industry participants in collaborative research. Otherwise, this is just a “feel-good” exercise, which the industry participants will soon see through.

As a summary, major recommendations are listed here *in prioritized order and as short-term (ST; 1 – 2 years) and long-term (LT; 1 – 5 yrs)*:

1. Fix the problem with developing a unique trip ID to integrate FDD. Try to apply some of the data management systems developed through the fishery independent data. (ST)
2. Prioritize sampling (both fishery independent and fishery dependent) with an eye to implementing recommendations of the 2009 ACL white paper re. partitioning of uncertainties to data sources and optimizing sample size. (LT)
3. Explore how to provide a dedicated vessel for shallower surveys to replace reliance on disparate state surveys. (ST for exploring options, LT for implementing)
4. Consider future data needs as FMCs move towards ecosystem approaches to management. Continue to explore other data collection methods for fishery-independent data (acoustics, mid-water trawls, HABCAM) as well as current survey design. (LT)
5. Continue to support data management efforts and expand successes to FDD. (LT)
6. Work to support more sea days for observer coverage. Use existing filtering process to allocate additional sea days. (LT)

Finally, some overall points for consideration for future reviews:

- Provide more directed guidance to the panelists before the meeting, rather than just reference to a website. A pre-meeting conference call could have worked through the terms of reference and given some more details on expectations.
- It would be helpful to have a “cheat sheet” to guide the panelists through the background documents. Some of the documents were entirely too detailed, and higher-level documents were needed. Panelists did not need many of the documents that were on the website.
- Linkages among the different aspects was somewhat lacking. An overall framework document would have helped to describe the regional landscape, the changing management regimes that the NEFSC is trying to respond to, and what general limitations are seen by the NEFSC staff in the fishery independent and fishery dependent data.
- Some of the presentations were entirely too detailed for the task at hand, and many were geared towards putting their piece of the whole enterprise in the most favorable light possible. This is totally understandable, but the “challenges” aspect would really have

been more useful to the panel. Some soul-searching by presenters and initial screening of the ppts by the Center personnel setting up the meeting would have helped in this regard. Presenters should be reminded of the panel's charge – not reviewing the value of every aspect of the programs presenting, only how their data were used and related to stock assessments.

- It would have helped this reviewer to start with the stock assessment presentations and follow with the data that went into the stock assessments. Similarly, presentation of the observer program overview before the details of pre-trip observer allocation and vessel selection would have made more sense.

## Northeast Fisheries Science Center Science Data Collection Program Review

Reviewer 3

### Fishery independent sampling

The presenters did a fine, detailed description of the bottom trawl, clam dredge, scallop dredge and shrimp trawl surveys. I found the survey design to be adequate in all cases and the lengths of the time series to be exceptional. Detailed comments for each survey are:

1. Bottom trawl survey. I found it difficult to understand why there is both a spring and fall survey, because there must be a tremendous degree of correlation between the two surveys and therefore data redundancy in the stock assessment modeling process. I know that once a time series has become as long as those of the bottom trawl surveys it is very hard to change, but perhaps the ship time devoted to one of the surveys could be better used, for example, to conduct the experimentation needed to estimate survey catchability.

The power of a time series to inform a stock assessment model depends upon the constancy of catchability as well as the series length. The one major event that altered the catchability of the bottom trawl surveys was the switch to the new vessel and trawl. Several of the presenters commented that the new combination produced larger catches, which implies higher catchability. The switch was accompanied by an amazing number of side-by-side calibration tows which should have produced very precise calibration functions to convert the catches of one vessel-trawl to those of the other. I believe, judging from both the initial presentations and the follow-up with Paul Rago, that the calibration work was sufficient to prevent any obvious catchability step in the time series, yet I still have doubts.

I had still other concerns about potential survey changes that are both abrupt (there was once an extensive Albatross-Delaware comparison, but I guess the Delaware was never used) or gradual (technology creep), but I guess all surveys have some of this and little can be done retrospectively. First, one presenter showed a plot from the herring time series and an apparent step in the abundance trend that was coincident with a change in the trawl doors, but I don't know if this case was the exception or if other time series for other species were similarly affected. Second, a new trawl performance monitoring system was described which better allows determination of whether a tow should be considered as acceptable or not. If really effective, such a system would better recognize and eliminate poor-performance tows. This is good from the perspective of standardizing survey performance but potentially bad from the perspective of a technology creep in survey catchability. At the Alaska Fisheries Science Center we agonize over such changes because of the duality of good and bad that accompanies them and the fact we can't afford the experiments needed to calibrate or even demonstrate the effects of small changes in methodology. Regardless of whether or not changes occurred, the principals of the national trawl survey protocols program still stand and we must guard against technological creep in survey catchability.

I am a bit concerned that the trawl survey fishing is done 24 hours, when there are many studies showing day-night differences in efficiency. All west coast surveys sample only during the day and several European surveys (Norwegian Barents sea survey; International bottom trawl survey) sample 24 hours but use model based estimators that correct for day-night differences. Although diurnal variability should only add uncertainty and not bias, perhaps a model based estimator (GAM with spatial covariates, time of day, spatial correlation, etc.) would be an effective way of reducing the variance of the survey estimators. Perhaps this is already done, but there was a lot of emphasis in the presentations about the stratified design so I assume that a design based estimator is being used.

Although there are many trawl performance variables that are measured on the surveys, it looks like there is no measure of swept area or the calculation of fish density within the tow path. I understand the problems associated with change, but calculating the relative abundance statistics in terms of swept area abundance estimates has advantages, especially if you intend to experimentally estimate survey catchability because the modeled population and the abundance index would be on the same scale.

I am also a bit concerned with the stratification scheme, because of the allocation of stations in proportion to stratum area. If there was only a single species, then you would have probably used Neyman allocation (station number set proportional to abundance or to variance), which reduces survey variance, but with the multispecies nature of the survey, it is difficult to justify any single species-based allocation scheme. However, from my experience with Alaskan bottom trawl surveys, stratification without variance allocation hardly reduces survey variance at all. In addition, since you have a minimum sample size per strata (3) you are actually oversampling the small strata which results in an overall reduction in any potential benefits of stratification. I recommend that you compare the current variance estimates to those ignoring any stratification to judge what increase in survey precision is actually gained by the stratification.

I was very impressed with the fish sampling scheme used on the survey vessel, it seems so smooth and efficient. I was also impressed with the flow of data from the FSCS collection scheme into the data base through the suite of QA/QC error checks. We are working on a similar system for Alaskan bottom trawl surveys, but large catches and variable vessel layout makes the Bigelow version impossible for us – but perhaps the portable version would work.

2. Scallop survey. I really like the idea of the habcam method of surveying the scallop stock, when combined with dredging to get biological samples. However this represents a big change in survey design and will require extensive calibration work with the dredge. I know there is concern about the low efficiency (0.40) of the dredge, especially by the fishing community who apparently believe that “missed” scallops results in an unnecessary reduction in their quota. I therefore recommend that you evaluate other options with the existing dredge survey to make sure it is truly broken before you replace it with something new. Perhaps use the habcam to estimate absolute density which can then be used to estimate the catchability of the scallop survey. Remember, the power of any survey to inform stock assessment models depends upon the length of the time series and the consistency in the survey catchability, therefore changing a survey is a big expensive deal. Make sure that there is a real uncertainty payback from the assessment modelers’ perspective.
3. Clam dredge. If you want to change the dredge and vessel then do classic side-by-side sampling to estimate a correction coefficient rather than determining the catchability of each using a depletion study to avoid the compounding of errors. If you can demonstrate that the depletion study can produce good estimates of sampling efficiency, then use them as priors on the catchability parameters in the assessment model. Remember that the sampling efficiency of the dredge likely varies with the texture and compaction of the bottom sediment, so you should repeat the depletion experiment at several sites that are representative of the various bottom types across the survey area, then calculate a survey mean estimate for the survey catchability estimate.
4. State surveys. I just don’t see the utility of including all these surveys in the stock assessment model because of the likely inter-series correlation and redundancy of information. If all were calibrated to the NMFS survey trawl, and if the efficiency of the NMFS survey were estimated using experiments, then the survey catches could be summed to estimate total population biomass. I wish more had been said about the coordination between all of these surveys both in timing, gear design and survey design. Unless some of the species being

assessed were really restricted to near shore areas, the NMFS survey likely has better spatial coverage of the fish stocks than the sum of the areas covered by the state surveys and should therefore be able to represent the entire stocks. However, this would not be true if there were substantial time variation in the proportion of the stocks occurring in state waters. In any event, I see the political motivation of including the state survey data, but I just don't see the scientific justification for doing so.

### **Fishery dependent data**

1. I found the description of the vessel log data, dealer data and the port sampling data very complicated and difficult to piece together coherently. It seems that random samples by weight are drawn from individual vessel landings that were presorted by size, although for some species the fish were already headed so length had to be estimated. To me it seems like the vessel logbook data is used to determine catch location, the dealer data is used to determine catch weight or numbers by species and the port sampling data is used to determine the length or age distribution of the catch and that the observer data is used to determine the discarded proportion of the catch. It was hard to understand how the sampling was portioned among ports, dealers and vessels. Perhaps the sampling is just fine, but the methodology was presented with such detail that I just lost the main issues and was left with the impression that the data collection system was overly complicated and that considerable effort was needed to piece it together to create the data really needed for stock assessment. I think that the whole data collection program could be redesigned to make it more efficient to produce the necessary data to inform the assessment models.
2. Sport fishing estimation. This is traditionally a very hard sampling problem because of the spatially diffuse nature of the catch and landings; however the description of the data collection was clear, logical and gave me the impression that the program was capable of producing sport catch estimates sufficiently good for assessment modeling. I suspect the sport catch is likely underestimated and always will be because the survey responders are likely not a truly a random sample of all fishers. However, I think this is definitely a quality program.
3. Observer programs. We heard three talks on observer programs. However, the last, which discussed the entire inter-related program, should have been first so that we realized that the first two talks were about distinct observer programs. The last speaker should be commended for logically bringing everything together in a more understandable package. It initially seemed that the "standardized by-catch reporting methodology" was the sampling program and the "vessel selection ..." was the methodology for getting observers on the vessels. I was totally surprised to later find out that they were two distinct programs. That said, I think the process of allocating the limited observer time among vessels, and areas was well thought out and likely produces the best (most informative) data for the cost. I expected to hear something about the efforts to develop video-based by-catch observing systems but there doesn't seem to be work on such systems in New England, which I find surprising because it is the way to go for small boat fisheries. I feel that more complete observer coverage might be a solution to the current complex way that fishery dependent data is now collected.
4. I am impressed with the cooperative research activities, except there doesn't seem to be a real focus on producing data products that are clearly informative for stock assessment. I agree that anything that supports cooperative activities with commercial and sport fishers is extremely important in the NE region because of the past history of antagonism. However, the products of any joint efforts should be traceable to their impact on the stock assessment process. Thus, for example, the details of any industry trawl survey should be worked out with the stock assessment folks beforehand to insure that the data is usable in stock

assessment models. Otherwise, the fishing community might consider that their efforts are not taken seriously by the scientific and fishery management community. I applaud the efforts of this program, but it seems a bit more focused on improving the relations with the fishing community rather than producing useful data.

## **Integration**

1. Fishery dependent data allocation process. I found the whole process of connecting the vessel log data, to the landings data (where biological sampling occurs) and assigning everything to the appropriate statistical area as overly complicated and seemingly required because the original data was not collected appropriately. One of the speakers pointed out many of the current faults and proposed some potential fixes that seem quite logical and appropriate. In a perfect world there would be 100% vessel coverage by observers so that all data, including biological sampling, effort measurement, catch and discards, could be collected on a tow by tow basis. Perhaps some of the necessary information could be obtained using video observing systems, but I think that the technology is not yet ready for use on the small and medium sized vessels in the New England fleet. I think that better electronic logbooks could also help, especially if the fishers estimate the catch weight by species by tow and enter it along with the effort data in the log. From what I have heard this would require better buy-in by some sectors of the fishing industry to become effective.

2. Data integration into the stock assessment models. I know that the use of several fishery independent time series with potentially conflicting time trends is more of an art than a science; still, I think the methodology and rationale for doing this needs to be closely examined and justified for its scientific contribution to a stock assessment rather than for any political reasons. If these indexes of relative abundance remain in assessment models then their relative weighting should be based on something other than the inverse of their variances because that ignores the uncertainty in the availability of the stocks to the individual survey and this could vary substantially from year to year. I understand that the data from each state's survey is essentially no-cost to NMFS and that there is political pressure from each state to use its survey results in the assessment model; but, I think that inter-series correlation results in data redundancy therefore each survey should have much less weight than a survey that encompasses the entire stock as does the NEFSC bottom trawl survey.

3. Although every sampling device has its own catchability, for weakly mobile species, habcam looks like it might be able to produce absolute density estimates that could be used to estimate the sampling efficiency of the NEFSC bottom trawl. Such sampling efficiency estimates could then be averaged over the survey area to produce an estimate of survey catchability that could, in turn, be used to set a prior on the survey catchability parameters in the stock assessment model. Just like trawls, there will be fish attraction and avoidance of habcam that this will have to be experimentally estimated before habcam estimates can be used for mobile species.

4. The RV Bigelow was specially designed to be an acoustics vessel, yet the primary acoustics survey is for herring and the data from this survey was not used in the herring assessment model because it was found to be inconsistent with the other model inputs. Although this was discussed well by both the stock assessment modeler and the acoustic assessment scientist, I think the issue really needs to be researched and resolved to everyone's satisfaction to build confidence in the utility of acoustic stock assessment, which I believe has the potential to be useful in a lot more situations than it is now used for.

5. The timeliness and quality control of both the fishery independent and dependent data seem just fine. I am especially impressed with the bottom trawl survey program and the sport fisher survey program. However the vessel logbook, and landings sampling programs need improvement.

6. One of the examples of the use of fishery dependent data was for the tilefish which is managed based on catch and effort data from a long line fishery. I understand that this is a small, localized, fishery which is not sufficiently large to justify a specialized survey. However, I think it might be possible to utilize the fishery itself, perhaps with funding from NCREP, to develop an experiment, perhaps using habcam, that might be able to calibrate fishery cpue to absolute density estimates. The problem with the current small fleet of vessels is that each vessel has its own cpue profile and when vessels drop out or are added to the pool, the mean is much more likely to reflect the vessel change rather than a population change.

### **Overall impressions**

I think that the fishery independent sampling programs of the NEFSC are world class, but that the fishery dependent sampling programs could be greatly improved. From my outsider perspective, perhaps one reason for this the continuing distrust of the fishery management community by the fishing community which results in the reluctance to fill out vessel logs, the misreporting of fishing locations, the unwillingness to carry fishery observers, etc. The NEFSC supports considerable outreach and joint fishing work, yet the lack of a truly cooperative approach to fisheries management apparently still remains. Perhaps this attitude will remain until the stocks rebuild, the number of unproductive fishers decline, and the level of income increases, but efforts to engage the fishing community need to continue or even increase so that fishers understand the importance of their data and see the value of their effort to fill out the reports, and host observers.



## Northeast Fisheries Science Center Science Data Collection Program Review

Reviewer 4

### Introduction

The objective of this report is to focus on a few key recommendations that might assist research and management of groundfish in New England. I have therefore resisted the inclination to comment on the many remarkable past achievements and current activities that are being conducted, with three exceptions.

I note the remarkable level of documentation related to surveys, sampling and catch data collection and management. These documents are very time consuming and difficult to write. Secondly, and in a similar vein, I noted the obvious passion and professional commitment of people to data collection and management. The relative importance of the permanent people infrastructure in database management is frequently dismissed in presentations by some IT people to senior management who argue that all data issues devolve to a one-time decision and cash infusion regarding the choice of hardware and software.

Finally, I compliment the staff for responding so positively to the review. My recommendations tend to follow the order of presentations rather than the terms of reference.

### Survey

1. ***I recommend that staff consider designing a meta-data history of idiosyncratic issues in the data collection.*** The idea is to capture and highlight the many and very minor issues like species creep (time varying species id) that fall below the radar screen of normal documentation yet can influence analyses if unbeknownst to numerical analysts. This is especially a risk if assessment staff carry serious assessment loads and little time to consider these issues and/or work remotely from those the people who have collected and archived the material. These risks increase significantly as such core databases are made available on-line.
2. I note that NEFSC staff have created a portable charter-friendly version of FSCS. If not already done, ***I recommend creating versions that can be used in all sampling applications including state port sampling.*** Apart from obvious advantages offered by electronic capture it would help propagate the use of the underlying data model of FSCS.

### Catch Monitoring

3. This recommendation relates to the issue of partial observer coverage. It appears that there is reason to assume that partial coverage may be associated with an observer effect and resulting biases in catch reporting. However, it is equally obvious that staff are coping as much as possible with these uncertainties. Considerable effort is spent exploring the data for signs of pathology and the assessment staff are attempting to cope with possible biases by examining how robust the model behaviour is to plausible catch biases.

We were also informed that there are ongoing attempt to make the program more efficient and explore alternative tools (i.e., electronic monitoring). However, it is much easier to estimate the costs of higher coverage rates than estimate the costs of not doing higher coverage. Notably among these is the potential lower revenue if quotas are reduced proportional to uncertainty, if the uncertainty relates to catch estimation.

Furthermore, in the ongoing B.C. experience, industry and the Department continue to find new ways to exploit the benefits of increased and full coverage that were not anticipated during design and implementation. Because of the examples provided below and others, ***I recommend that***

***discussions of the cost and benefits of higher levels of coverage not be closed and efforts continue to assess the real costs of not committing to higher coverage (or even benefits of lower coverage).***

It is worth noting that there is a “tipping point” with respect to some of these benefits in that some can only be realized at coverages of effectively 100% as opposed to even 90%.

With respect to value-added issues in the fleets:

- It was commented that there are situations wherein fishers probably discard marketable specimens because the quota for that species is attached to another sector. In the B.C. context, the transition to full coverage with ITQ resulted in no legal necessity for a harvester to discard marketable specimens of any species, provided they had quota. Some fishers have commented that the extra revenue generated from this change surpassed the cost of enhanced monitoring.
- Higher levels of monitoring may assist in gaining access to or staying in specific markets, or higher prices. This could be as simple as making a difference in applying for eco-certification. In B.C., the trawl fishery, as part of a program to obtain eco-labelling and thereby retain a lucrative market beyond 2014, partnered with ENGO's to develop a more eco-friendly bottom trawl footprint. This involves a combination of both freezing and shrinking of the historical footprint as well as an aggressive coral encounter protocol wherein known coral sites were excluded from within the permitted footprint with fairly severe individual catch limits. Furthermore, as new spots are encountered (and noted by observers) these are added to the excluded location within the allowed footprint as soon as reported by the observer.

Without getting into more examples, the point is that the lesson being learned in B.C. is that long after the introduction of full coverage, managers, scientists and enforcement groups are still finding significant new opportunities for cost-saving and/or sophisticated management devices to solve new issues and that many of these devices are only possible with accurate catch data at the level of each trip.

Finally, while no cost/benefit analysis in the NE context may ever indicate that the overall fishery should receive higher or full coverage, ongoing consideration of the issues these may reveal that it would be appropriate for individual sectors.

#### **Stock Assessment**

4. In my opinion, it appears from the presentations and question period that the NEFSC is not currently investing sufficient resources into the simulation approach offered by Management Strategy Evaluation (MSE) (i.e., the work of de la Mare, Butterworth, Cox and Kronlund among many others). ***I strongly recommend committing significant resources to MSE.*** It appears to offer major benefits to the NEFSC with respect to, among others:

- An analytical approach which focuses on developing effective advice tested with simulation to meet fishery objectives rather than the traditional assessment advice which focuses on improved precision and reduced bias in estimated current stock size with forecasting.
- An approach that provides a means for quantifying the impact (value) of the varying data inputs, for example testing the impact of reduced survey frequency or intensity on meeting management objectives, as opposed to variance targets. These activities offer the potential to provide a basis to significantly reduce or re-allocate funds.
- An approach that provides an effective vehicle for more meaningful cooperative research and stock assessment with the fishing industry, particularly in the objective definition phase.

5. With respect to all the numerous “minor” species that are not on the assessment radar screen, ***I recommend developing a minor species status screening tool.*** This tool will review

available data for signs of decline. Screened species can then be subjected to more intensive review and possibly assessments if deemed necessary. The enhanced early detection might provide the opportunity to implement mitigation measures before more drastic measures are mandated owing to EPA-legislation. The screening can be as simple or complex as is deemed necessary but should be updated at regular intervals.

#### **Enhanced cooperation with industry.**

6. Numerous comments were raised about outreach to the fishing community concerning credibility of the NEFSC surveys and assessments. There was mention of numerous meetings to which industry is invited. In my opinion, much more is accomplished in smaller scale, less official settings, particular one-on-one. ***I recommend that every attempt should be made to fully embed industry individuals in the assessment process and in turn those individuals should be expected to assume some responsibility for at least some elements of the assessment.*** An effort should be made to search out individuals who have been critical of previous assessment. It is understood that this different relationship will take time and probably numerous attempts to mature.
7. ***I also recommend that efforts be made to develop an NEFSC-Industry position to serve a liaison role with a focus on assessment and survey sampling design.*** This person, who needs to have the trust of the industry, would focus as a translation vector on these issues. There are numerous examples in the Alaskan and B.C. context of these individual helping bridge the divide between industry, particularly on assessment and survey issues.

#### **Suggestions for the next review in for 5 years**

8. I recommend the following changes with respect to the next review:
  - The panel be provided with brief NE fishery overviews (i.e., 5-10 pages) and that the meeting begin with a overview on the nature of the NE fishery context.
  - The panel be provided with previous reviews or strategic document that pertain to the same issues. It seemed like material and presentations were more geared to whether the whether specific tasks were done right, rather than whether the right tasks were being addressed.
  - That a more concerted effort be made to bring in external people who have expressed dissatisfaction with current research efforts. These should include various interested parties, including processors, ENGOs, harvesters, and state representatives.

## Northeast Fisheries Science Center Science Data Collection Program Review

Reviewer 5

1. *To what extent do fishery independent survey data quality, statistical precision, and timeliness issues impact overall assessment accuracy, precision and timeliness?*

It was difficult to address this item completely, and certainly not quantitatively, because the links between the various sources of survey data, their precision, and the associated assessments was not demonstrated. In the few example assessments that were presented, it was not always clear what the quality of the survey index was, and the links between survey precision and the accuracy of assessments were not demonstrated. This is generally difficult to do in any case, but it does highlight a drawback of the diagnostics of the stock assessments. Some consideration should be given to developing assessment methods which evaluate the contribution of each of the major data sources (and see point 4, below, for examples).

Fishery independent data quality in terms of the collection, archive, and delivery to stock assessment staff, is exceptionally good and very timely. Survey accuracy (bias and precision) was more difficult to evaluate. Some evaluation of survey index performance needs to be regularly reported and survey scientists should be encouraged to do this as part of the QA of the survey data. These would include plots of the survey index time series with 95% confidence intervals (some of these were shown); internal consistency plots (plot of age  $n$  in one year against age  $n+1$  in the following year for each age); catch (cohort) curves of the (log) indices at age; and segmented post plots of the indices at age (pie charts centred on sample locations where the size of the pie is proportional to density and the size of the segments in the pie are proportional to age proportions). An

estimate of  $Z$  for the age classes for which  $F$  is reported in the assessment might be useful too, given that this would be so easily derived from a linear fit to the catch curves. It was mentioned that these may be produced for some stocks as part of the data screening workshop prior to the assessment meeting, so they may already be available for some stocks.

2. *What are the major fishery independent survey successes and how should they be supported?*

The groundfish survey is a clear success. The long time series, the vast number of species covered, robust statistical design, large geographical coverage, and technical facilities available (FSV Bigelow and its suite of modern equipment) are factors which stand out. The attention paid to inter-vessel calibration was exceptional, both in terms of the effort applied at sea (specifically the huge number of paired tows) and in the analysis. Support for both surveys should continue, if nothing else because the data generated supports the biological and ecological science programmes, as well as numerous groundfish stock assessments. It might however, be instructive clarify the merits of each of the two groundfish surveys (i.e. which have the most impact on which assessments).

The Bigelow seems like an excellent research vessel. The new electronic fish measuring system on board the FSV Bigelow is absolutely world class and this integrates well with a comprehensive database. This is being made use of effectively; however, it is unclear where the efficiencies gained from the system have been made. The gear mensuration system is world class and looks to provide accurate swept areas. These should be exploited more effectively by

delivering minimum biomass estimates from the surveys for all species.

Habcam is a superb system for visual surveys. This should be supported and developed, particularly given the value of the scallop fishery. There is potential to use this for tilefish, anglerfish, skates and rays, and in areas where trawl cannot be deployed. There should be support for analysing the data automatically by developing image analysis systems to deal with the large volume of images which will be generated.

The expertise available in fisheries acoustics, both in house and locally (WHOI), is world class and should be made more of. The analysis of acoustic data is excellent and there is some good research carried out in support of the activity. Training should be provided to additional scientists to process acoustic data to reduce the risk of loss of expertise.

The fish diet database, based on 600,000 stomachs, is world leading and plans for the effective deployment of analyses derived from it in assessments are good, but these plans need to be shored up and prioritised. Given the increasing size of natural mortality relative to fishing mortality it would seem that an immediate priority would be to develop size/age based estimates of natural mortality from these data for the most important assessments.

*3. What are the major fishery independent survey limitations/weaknesses and how could they be resolved? Define potential improvements and priorities for recommended improvements.*

No major weaknesses in the bottom trawl survey are apparent. More could be made of historic tow locations to reduce time spent searching for suitable ground on the bottom trawl survey. Some effort to determine whole gear selectivity could be started to achieve a long term objective of estimating absolute abundances, particularly for those species for which there is no catch data. Minimum biomass estimates (based on the area swept by the doors) could be reported

for the latter in the meantime. With the exception of tow location, these are probably low priorities given current resource constraints, but should be considered in the medium to long term.

Scallops are by far the most important resource in terms of landed value. Given the selectivity of dredges it would seem prudent to continue to develop Habcam as a complementary survey tool (and much more could be made of this tool, see above). Visual techniques in the deep sea are developing world-wide both for more sessile organisms such as shellfish and coral (Roberts et al, 2005), as well as fish (McIntyre et al., 2013), and NEFSC has potential to be a world leader with the Habcam system. Links to WHOI ([www.whoi.edu/main/auvs](http://www.whoi.edu/main/auvs)) should be enhanced to exploit their expertise in Autonomous Underwater Vehicles (AUVs) as a potential platform for the Habcam system (Yoerger et al., 2007). In the long term this would introduce massive savings on research vessel costs. If seabed type is important to the scallop survey, split beam echosounders could be employed to assess seabed type in the absence of more expensive multibeam systems. Methods for determining seabed type using scientific echosounder systems have been developed recently by the Advanced Survey Technologies Group of the South West Fisheries Science Centre.

The acoustic survey time series of herring biomass estimates has not provided a good fit to the herring assessment since 2004. Reasons for this remain unexplained but are likely due to restricted survey coverage. Acoustic survey coverage should be more extensive, albeit stratified to cover Georges Bank more intensively, and a suitable time to survey should be determined. At the time of the surveys done to date, herring could have been in transit to any of the potential spawning areas around Georges Bank and the Gulf of Maine. There is evidence in some years of the acoustic surveys that fish were detected on the limits of the survey indicating that the distribution was not

contained. There is also the vast area to the south which has herring throughout the area in the spring (Overholtz et al. 2004). Examinations of the acoustic data collected during both the spring and fall bottom trawl surveys will help determine when best to survey in future, but the entire continental shelf should be covered as part of a wider pelagic ecosystem survey. It would be prudent for a research institute as significant as the NEFSC to avoid having such a huge gap in the monitoring of a major component of the marine ecosystem. Acoustic surveys provide the only means to estimate the abundance and distribution of important pelagic species such as herring, sandeel, and mackerel. These species are important not only for stock assessments, but also for ecosystem considerations as forage species for higher trophic levels in the marine ecosystem. The method can also be applied to semi-pelagic fish species such as hake, butter fish and squid, as well as important micronekton such as krill.

The clam survey seems like the only effective source of fishery independent data for the two species. Although the survey is expensive, the resource is valuable so it would seem prudent to persist with this. Given the longevity of the species, the frequency of the surveys could be reduced to bi-annual events as has been the case in the past. The feasibility of this could be evaluated with a management strategy evaluation.

The northern shrimp survey has a quarter of its samples allocated as fixed stations. The documentation provided does not explain why this is. Unless there is a reasonable justification for keeping these samples in the design they should be re-allocated to the random samples.

The lack of co-ordination between the federal and state surveys represents a missed opportunity. These could be integrated more effectively to provide a synoptic comprehensive coverage of the inshore and offshore marine environments. This is particularly the case given that the new

federal vessel is larger with limited access to near shore environments.

#### *4. To what extent do fishery dependent data quality, statistical precision, and timeliness issues impact overall assessment accuracy, precision and timeliness?*

This is difficult to determine (as per point 1 above). There are few examples of assessment model diagnostics which apportion the uncertainty of the estimates amongst the various data sources. One nice example is the latest North Sea herring assessment based on a state-space assessment model. The report of this assessment presents observation variances for each source of data (see Figure 2.6.1.26 on page 243 of ICES HAWG report 2013<sup>1</sup>). The latter analysis is clearly not appropriate here, but if this TOR is to be answered in the future, some sensitivity analyses, or analyses similar to those presented in HAWG, would be needed. This will need some development of stock assessment models.

It is clear that heroic efforts go into examining the data quality of the fishery dependent data. However, the precision and accuracy may be compromised by the ad hoc nature of the sampling process. My colleagues will elaborate on this, but I would also advocate changing the scheme towards a probability based port sampling scheme for sampling the catch. The difficulties of integrating databases reinforce the need for a common trip definition: this problem should be solved as an absolute priority.

Financial audits should be considered to check the integrity of the landings information and ensure that dealers are not colluding with vessel skippers to under report landings. Catch limits provide an incentive to under report and in Europe this problem was only solved by following the money. Trends in discards could provide evidence for this: if discard rates did not go

1 Available at:  
<http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/HAWG/HAWG%202013.pdf>

up when catch limits were introduced, then either fishermen were either suddenly very good at avoiding fish for which they had no quota (in which case it should be determined how they do this), or they may be underreporting their landings.

5. *What are the major fishery dependent data sources successes and how should they be supported?*

The observer programme is extensive and well designed, with a lot of attention given to the detail of trip allocation. The prospect of observer bias is always an issue when coverage is less than 100%, so some examination of observer bias should be undertaken. One simple example is the comparison of the age/length compositions of observed trips with similar trips (in time and space). As with many observer programmes it seems that observers are difficult to retain. Providing observers with more diverse duties might maintain their interest and train them for future roles at the Centre.

The recreational fishery survey looks to be world class. Results from this should be displayed effectively online given the large number of anglers interested in this activity.

The biology programme is extensive and suitable for supporting both the fishery dependent and independent programmes. Many fish are aged (perhaps too many) and the principle of only ageing fish which have been validated is admirable. There are some key species where biological parameters need further investigation and research efforts in these areas should be supported. An obvious example is monkfish. This is a valuable resource and yet many fundamental biological questions remain which are relevant to stock assessment (growth, ageing, natural mortality, what happens to the older males? etc.).

6. *What are the major fishery dependent data limitations/weaknesses and how could they be resolved? Define potential improvements and priorities for recommended improvements.*

A single trip definition across all systems is lacking and should be determined for the most important systems at least (dealer and trip

reports). Port sampling design should be improved (see colleague's report).

In the age reading programme there seems to be little contingency for the loss of expertise given that each species is only covered (read) by one age reader. Age reading is a skill which requires substantial investments in time and training, so the Centre should have contingency and succession plans for the replacement of age readers: a second reader for each species would be advisable. This will require an internal programme of consistency checks and internal exchanges, but this would be good practice in any event.

7. *What recommendations do you have for prioritizing fishery-independent and fishery-dependent data collection improvements?*

Time for research should be ring-fenced for staff engaged in stock assessments. This should be no less than 20%. This will not only have long term benefits in developing better methods, but also reduce the risk of staff dissatisfaction and ultimately departure.

The relevant organizations should compromise on a unique trip identifier associated with a common definition of a fishing trip.

A systematic pelagic ecosystem (acoustic) survey should be designed with common objective of sampling herring as well as other components of the pelagic ecosystem.

The port sampling survey should be re-designed in accordance with the appropriate probability based on principles set out in ICES (2013).

Management Strategy Evaluation (MSE) techniques should be developed and applied to: (i) examine sensitivities of the data to management through existing Harvest Control Rules (HCR); and (ii) investigate HCR based on simpler stock assessments (survey based metrics?). Some consideration should be given to linking to existing methods (e.g. in Europe with FLR).

Habcam should be developed and applied to determine an index (or absolute estimate) of scallop abundance as well as other species in the

benthic ecosystem.

8. *To what extent are fishery independent and fishery dependent data readily accessible to Center stock assessment scientists and to various external researchers who may wish to replicate NMFS stock assessments?*

This was not clear. The survey data are not yet in a format for public dissemination. NEFSC should seriously consider using the DATRAS format for this (see <http://datras.ices.dk/Home/Default.aspx>). This would save them having endless debates about standards and formats and it would also allow European scientists to access their data with code that is already tailored for DATRAS type data.

Stock assessment summaries should also be provided online much like ICES (see <http://info.ices.dk/datacentre/StdGraPhDB/FishStockDB.mdb>).

The data available in stock assessment reports should also be available online in a form that allows for replication of the analyses.

9. *Identify the highest priority needs for improving fishery dependent and fishery independent data. Define potential improvements.*

See 7 above, but the highest priorities would be:

1. A unique trip identifier.
2. 20% research time for stock assessment staff.
3. Development of MSE to determine relative importance of data sources in management.
4. An improved port sampling design.
5. A pelagic ecosystem survey based on acoustics.
6. A benthic ecosystem survey based on Habcam.

## References

ICES (2013). Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 6 - 9 November 2012, ICES Copenhagen. ICES CM 2012 / ACOM:52 71 pp.

McIntyre, F.D., Collie, N., Stewart, M., Scala, L. and Fernandes, P.G. (2013). A visual survey technique for deep-water fishes: estimating anglerfish *Lophius spp.* abundance in closed areas. *Journal of Fish Biology*.

Overholtz, W., Jacobson, L., Melvin, G., Cieri, M., Power, M., Libby, D. and Clark, K. (2004). Stock assessment of the Gulf of Maine–Georges Bank Atlantic herring complex, 2003. *Northeast Fish Sci Cent Ref Doc*: 04-06.

Roberts, J.M., Peppe, O.C., Dodds, L.A., Mercer, D.J., Thomson, W.T., Gage, J.D. and Meldrum, D.T. (2005). Monitoring environmental variability around cold-water coral reefs: the use of a benthic photolander and the potential of seafloor observatories. In. *Cold-Water Corals and Ecosystems*, Springer: 483-502.

Yoerger, D.R., Jakuba, M., Bradley, A.M. and Bingham, B. (2007). Techniques for Deep Sea Near Bottom Survey Using an Autonomous Underwater Vehicle. *The International Journal of Robotics Research* 26(1): 41-



## **Northeast Fisheries Science Center Science Data Collection Program Review**

Reviewer 6

### **Background**

The National Marine Fisheries Service's Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts, conducted an external review from August 5-9, 2013, to evaluate its current scientific data gathering and management procedures. Specifically, the review focused on fishery-independent data, fishery-dependent data, biological data, and data management as they relate to fishery stock assessments conducted by the in the NEFSC. The review was conducted over a 5-day period. NEFSC staff (from the Woods Hole Laboratory and regional offices) provided presentations to a 6-member reviewer panel. The presentations described the NEFSC and cooperative research data collection and management activities and outlined procedural strengths, challenges, and opportunities related to each activity. NEFSC staff, partners, constituents, and the public were sitting in during the presentations, and were given the opportunity to ask questions at the end of each day. To supplement the on-site presentations, the reviewers were provided web-based access to extensive supplemental reports that described all aspects of the data collections and data management in detail.

This review report provides a brief assessment of the strength and weaknesses of the scientific data collection programs, and some recommendations for potential improvements in some sampling programs that NEFSC may consider. I will mainly focus on aspects of the survey sampling programs, and defer to other members of the review panel to provide advice on the development of management strategies to prioritize research and data collection efforts among programs.

### **Some general comments and impressions**

Clearly, the very experienced and highly skilled scientific and technical staff at NEFSC will have the best overview of the strength and challenges in their data collection programs, and in particular the administrative, budgetary, and logistical constraints for various alternative approaches. The heavy-workload related to data collections, management, and analysis to support in excess of 50 stock assessments seems to be so demanding that NEFSC will not have sufficient time to develop and implement solutions to some of the challenges they have identified. The timing (are assessments required yearly) and also quality requirements, and therefore data-needs for each stock assessment would ideally be prioritized to carve out time for research.

The link between stock assessment work and ecosystem studies at NEFSC was not explained in much detail during the review presentations. To develop an ecosystem-approach to fisheries management, it is important that stock assessment work be closely linked to ecosystem research efforts at NEFSC.

#### **I. Fishery-dependent surveys**

NEFSC have supreme expertise in the design and execution of fishery-dependent surveys, and very strong analytical capabilities. All fisheries-independent surveys run by NEFSC considered are statistically sound and have very high standards for quality assurance and quality control (QA/QC). Improvements in accuracy (precision and bias) in these programs will generally

require increased sample sizes, increased spatial coverage, and advanced experimental studies for determining gear efficiency. All of these improvements, which NEFSC clearly are aware of, will come at a high cost, and therefore it is important that data needs and level of precision required for stock assessments be prioritized. It is possible that data collections for some less valuable species could be reduced to allow more effort for other species.

The fishery-independent bottom trawl surveys run by NEFSC are designed according to sound statistical principles, and likely provides the longest probability-based time-series of fish abundance indices in the world. The quality assurance for standardizing the field sampling, and for coding and tracking biological data (e.g., the use of bar codes to label age samples and link them to a station) from sample collections to database storage is world-class and commendable. The well executed vessel calibration studies conducted before the switch to R/V Biggelow in 2009, and the corrections applied to the time series appears to minimize the effect of vessel change. The QA/QC procedures to standardize hauls are strong, but I would recommend that abundance indices by tow (catch per 20 min), or by area-swept (catch by x sqm), be corrected for actual towing time or for distance towed on the bottom. NEFSC has developed very good manuals and protocols for biological sampling of catches. The very extensive stomach sampling for diet studies provides high quality information for ecosystem studies, and for quantifying predation mortality for many species.

The age collections conducted by NEFSC are very extensive, and costly. It is likely that the number of fish aged by length bin can be reduced for some species, without much loss if any, in precision in estimates of age-length distributions. The reason is that most of the variation in estimates of numbers-at-age often can be attributed to the number of primary sampling units (e.g., trawl stations). The sources of variation, and the strategy for age sampling can be assessed through simulations based on historic data. I recommend that NEFSC conduct analysis to evaluate precision in estimates of abundance indices by age as a function of sub sampling sizes for age. For example, extensive analysis of data from the winter survey in the Barents sea conducted by Institute of Marine research in Norway show that 1 age sample per 5 cm length class suffices for Northeast Arctic Cod, and that virtually no gain in precision is achieved by sampling up to 10 otoliths per 5 cm length class (Aanes and Vølstad in prep, to be presented at the ICES ASC 2013). It is likely that NEFSC can increase the length bin from 1 cm to 5 cm for some species, and only sample 1 otolith per length bin. This could be evaluated through simulations (bootstrapping of tows, and resampling within tows) where bin-size is set to for example 2 cm and 5 cm. Otoliths can be chosen random from the otoliths collected per 1 cm for each bootstrap sample.

Another recommendation is that some more efforts be dedicated to assess age-reading errors, for example through cross-reader comparisons with other laboratories. The time for such efforts may be facilitated if the number of age samples collected can be reduced.

Another general recommendation is that diagnostics on sample sizes for length and age not only list the number of fish measured, but also include the number of primary sampling units the length and age samples were based on. In general, the effective sample sizes for estimating length- and age-compositions will depend more on the number of PSUs than on the number of fish measured.

For the bottom trawl surveys, another cost saving may be to reduce the time used to search for trawable area within a chosen "plot" by using historic information on executed tows. NEFSC spends significant time searching for a trawable spot within a 1-mile or 3-mile circle. This seems inefficient.

From the presentations, it appears that the new sampling trawl employed onboard R/V Bigelow often catches a large amount of fish in a standard 20 min haul. For large catches, sampling errors due to subsampling may offset the value of a larger sample. One recommendation is that NEFSC embed an experiment to build up data for choosing optimal towing duration. For, say, 5% - 10% of stations, picked at random, tows of varying duration could be taken (say 10 min, 15 min and 20 min) in random order. Over time this would build experimental data to determine optimal tow duration (trade-off between number of stations and time at a station). Such embedded experiments also would be effective for selectivity and catching efficiency studies since it would cover varying bottom types and a wide number of species over time.

For the acoustic surveys I defer to the two acoustic specialists experts on the review team for comments and recommendations.

The two surveys of scallops with HabCam and dredge provide very accurate estimates of scallop abundance, and would ideally be maintained due to the large economic value of the scallop fishery. One recommendation would be to use HabCam data to determine the optimum towing time for the dredge survey. This could possibly make the dredge survey more cost-effective. If the monitoring budget doesn't support the cost of the scallop survey on top of all other activities, an option could be to reduce survey efforts for less valuable species.

The surf clam/Ocean Quahog surveys provide reliable abundance indices and supports stock assessments for very valuable fisheries. The surveys are efficiently designed and executed.

The shrimp survey also follows a probabilistic design, augmented by 20 fixed stations. The utility and value of the fixed stations was not clearly documented. NEFSC identified some possible improvements of the survey related to the monitoring of gear performance and standardization.

State trawl surveys are used extensively in summer flounder stock assessment (12 state surveys with age sampling along the range of the stock.) The value of the state surveys as data-sources for stock assessments of summer flounder is dubious due to differences in vessel/gear catching efficiency. If the state surveys are included, they should be assigned to represent separate strata and weighted accordingly, and not be included on equal terms with the NEFSC surveys which have much larger spatial coverage. An alternative approach is to use the state surveys in a stratified estimate weighted by area as an index of abundance for the inshore areas. By following trends in this index, compared to trends in the NEFSC abundance indices, this could provide information on shifts in distribution between inshore and offshore areas.

Industry-surveys are also problematic to incorporate in stock assessments when stations are selected ad-hoc, and gears are different than the scientific sampling gear used by NEFSC. Collaboration with the industry would perhaps inform science more if efforts were dedicated to special experiments. In particular, sampling outside the spatial area covered in the NEFSC surveys may inform the evaluation of survey coverage.

## **II. Fishery-dependent surveys**

### **II.1. Vessel and dealer data**

The NEFSC staff identified several shortcomings of the current data from dealer and vessel-trip reports (VTR). I fully agree with NEFSC staff that the development and implementation of a unique trip identifier is the highest priority. A unique trip-ID would allow NEFSC to use many of

the very sophisticated QA/QC procedures developed for fishery-independent data collections (e.g., the tracing of samples using bar-codes) also in the port-sampling program.

Recognizing that vessel-trips may cross-statistical areas and stock boundaries, there are uncertainties in catch allocation at the high spatial resolution used in the management of some species. The accuracy of area-allocation could be improved through a unique trip ID and mandatory electronic logbook data (linked with VMS data when possible). This could also improve the timeliness of getting VTR data.

## **II.2. Port-sampling**

For at-shore sampling, the sampling of catches is taken from vessels and trips that can be accessed in ports where they land their catches. In these cases, a sampling-frame will ideally be based on a list of access-sites crossed with time (for example port-days). The primary sampling units (PSUs) can be defined as port-days, which can be randomly selected within a quarter and region. Within a PSU, a sample of trips (secondary sampling units, SSUs) can be taken, and the catch of a selected trip can be sampled by market-category, so fish for length measurements and ageing are collected in stages 3 and 4. There is one extra level of clustering in port-sampling surveys, as compared to at-sea sampling surveys (e.g., observer programs), where vessel-trips are selected in the first stage. This very high level of clustering suggests that it is not viable to get a direct simple random sample of fish from a trip.

The port-sampling survey conducted by NEFSC has some ad-hoc components that may introduce bias in estimates of catch composition of unknown magnitude and direction. Of particular concern is that the selection of port-days and vessel trips for catch sampling appears to be controlled by port samplers, and not according to a design that can assure representative sampling over time. In the analysis to estimate length composition and numbers at-age of catches based on data from the port-sampling program, NEFSC treats the vessel-trips as primary sampling units, and it is also implicitly assumed that catch samples for each market category are obtained from a simple random sample of trips from each domain (e.g., stock-area, fleet, gear-type). For this assumption to be reasonably supported, the port-visits and sampling of trips within selected ports would have to be scheduled so that vessel-trips across ports have similar (ideally equal) chance of being selected within each "stratum". The sampling "strata" used by NEFSC is based on region, species, market-category, gear, and stock-area. In actuality, for this level of grouping the actual number of trips is not known in advance of a sampling event, and the sample size at the trip level cannot be controlled for.

The yearly biological sampling requirements specifies the number of length and age samples per species and market-category for each region, and not the number of site-visits, and number of trips. This procedure implies a quota sampling, where the port-samplers fill the sample requirements but do not necessarily spread out the sampling across port over time to achieve a representative sample of trips. The information provided on sample sizes for length and age generally only specified the number of fish measured, and not the number of trips sampled. Due to intra-cluster correlation, the precision in estimates of catch composition by length and age will typically be driven by the number of primary sampling units, and to a lesser extent by the number of fish sampled from each PSU.

It is recommended that NEFSC move towards a more rigorous statistical approach to port sampling. A survey sampling method that better controls the selection of port-days and trips within port-days reduces the risk for bias in estimates of length and age compositions. Since estimates of catch in numbers by length and age are a key data source in analytical stock-assessments, it is important that sources of bias be eliminated to the extent possible within logistical and financial constraints. In the ICES community, a series of workshops since 2007

have dealt with practical implementation of statistical catches sampling programs (ICES WKPRECISE, WKMERGE, WKPICS1, WKPICS2). ICES will likely extend these efforts by establishing a long-term working group (WGCATCH) in 2014. The objective is to move away from the métier-based quota-based sampling approach that has been widely applied in the EU, and towards probability-based sampling where sampling frames and multi-stage sample selections are clearly specified. A key element is to employ stratified probability-based sampling which has the advantage that sample sizes and sample selections at the primary level (PSUs) in each stratum can be controlled, in advance, thus minimizing the need for imputations to fill in data gaps. Design-based estimators will allow samples to be easily extrapolated to the target population of catches using weighting factors based on inclusion-probabilities. Estimates for domains such as catch of a stock within an area and gear type may be based on post-stratification or other re-weighting procedures. Target sample sizes by domains (e.g., by region-species-market-category-gear-stock-area, and quarter) may be reasonably controlled for by using historic landings data for each port in the sampling frame. The Scottish case study presented in ICES WKPICS2 provides an example of a probability-based port-sampling scheme to fill catch data needs similar to those described by NEFSC for Northeast US fisheries. As a co-chair of the upcoming ICES WKPICS3 meeting to be held in Copenhagen, Denmark, during November 2013 I would welcome the participation of NEFSC staff. I also recommend that NEFSC explore the possibility of involving the consultants that worked with NOAA on the Marine Recreational Information Program (MRIP) to help develop a more rigorous statistical approach to port sampling.

### **II.3. Marine Recreational Information Program (MRIP)**

MRIP is a state-of-the-art probability-based survey to estimate catches in the U.S. marine recreational fisheries. The development of this survey is an example of successful partnership between NOAA and world-class survey statisticians from US industry and academia. It has been proven that it is possible to implement a sophisticated design and estimation methods to quantify catch (harvest and fish released) in recreational fisheries. Issues with reliability of self-reported catch-and-release and discard of fish by species remains, but likely cannot be much improved within reasonable cost. Increased precision in catches estimates for species where recreational catches accounts for a significant portion of the total catch (e.g., summer flounder) can mainly be achieved by increased sample sizes in NE region (using MRIP methods).

Shortcomings of the data from the recreational fishery that cannot easily be resolved are its coarse spatial resolution and the lack of biological samples (length, weight, and especially hard parts for aging), as well as the uncertainty in self-reported catch-and-release or discard estimates (e.g., recall bias), and complete lack of biological data for the discards. For species where recreational landed catches accounts for a large portion of total removals in a fishery, such as striped bass (74%), summer flounder (40%), black sea bass (59%), Gulf of Maine cod (24%) and bluefish (70%), clearly there will be potential for bias in estimates of length age distributions of catches based on data from other sources. There are no easy solutions to correct such biases, and effects on stock assessments may best be assessed through sensitivity analysis.

### **II.4. Observer programs**

The design and operation of the observer-sampling program is impressive. NEFSC has developed an effective system that allocates observer coverage to fleets where discard is likely to be substantial and of importance to the reliability of the stock assessments. The randomization of the vessel selection supports design-based estimates of discards, for example based on the combined ratio estimator. The automated system for trip selection and employment of observers implemented in 2010 is cost-effective and minimize bias in trip selections. The one issue with

observer programs in general that is difficult to assess is the observer-effect on data-collections (changes in fishing behavior when observers are onboard). If this effect were large, then 100% coverage would be the only way to eliminate biases in estimated discard rates by spatial areas and species. The very complex management regime in the North East may suggest that discard levels, fishing within closed areas, etc. can be affected when observers are onboard. NEFSC use a number of methods to diagnose sources of bias (check trip length of observed vs. unobserved trips; check kept pounds per trip for observed vs. unobserved trips) that follows best scientific practice.

### **III.1. Integration**

A very large challenge for NEFSC is to provide assessments for a large number of stocks at the fine spatial scales required to support changing management demands. The spatial strata in the fisheries-dependent surveys, and the vessel trips sampled in the fisheries-dependent surveys crosses stock-area boundaries. Sample sizes within management units may be insufficient to provide reliable estimates of stock size and the composition of catches. In response to management requirements, sampling of commercial catches is set up to chase data at métier level. It is well documented in the literature that quota sampling can bias the overall estimates of catch at length and age. A change in sampling plans from year to year, and the extensive data management efforts to support management comes at a high cost. These tasks takes much resources, and NEFSC has other research needs to cover than groundfish stock assessment.

It is recommended that the inclusion of data sources in stock assessments be based on scientific, and transparent criteria. More data will not necessarily improve the accuracy of stock assessments. For example, the inclusion of state surveys that are conducted in-shore on equal terms with NEFSC surveys with extensive spatial coverage may introduce biases of unknown magnitude and direction.

### **Final recommendations**

1. It is recommended that NEFSC moves towards a more rigorous statistical approach to port-sampling. More engagement in ICES expert groups, and the involvement of survey statisticians from Industry and academia could help NEFSC in this effort. This would not only help reduce biases in estimates of catch composition, but also allow more accurate estimates of the precision in estimates relative to sample sizes at the various stages of sampling (from ports, trips, to fish within market-category). This way, the appropriate level of catch sampling can easier be determined.
2. The prioritization of sampling programs and the choice of assessment models should not only take into account the value of commercial fisheries they support, but also data needs for ecosystem studies that will support the development of ecosystem-based management over time. This may be based on managements strategy evaluations.
3. NEFS has a very large effort in age collections. It is recommended that sampling strategies for age-collections in the fisheries-independent surveys are evaluated to assess if age collections can be reduced to save cost.

## **Appendix I. Terms of Reference (TOR) for 2013 Data Collections Science Program Reviews**

### Objective

The objective for these reviews is to review and evaluate the Center's current scientific fishery-dependent and fishery-independent data as it relates to fishery stock assessments conducted pursuant to the Magnuson-Stevens Act:

- NOAA ship-based surveys
- Cooperative research surveys
- Logbook and observer data
- Data management and quality control

Reviewers will provide advice to the Center on the direction and quality of these data collection and management programs

Using as context, two-three or more typical and important stock assessments conducted by the Center, reviewers should address:

### ***TERMS OF REFERENCE***

1. To what extent do fishery independent survey data quality, statistical precision, and timeliness issues impact overall assessment accuracy, precision and timeliness?
2. What are the major fishery independent survey successes and how should they be supported?
3. What are the major fishery independent survey limitations/weaknesses and how could they be resolved? Define potential improvements and priorities for recommended improvements.
4. To what extent do fishery dependent data quality, statistical precision, and timeliness issues impact overall assessment accuracy, precision and timeliness?
5. What are the major fishery dependent data sources successes and how should they be supported?
6. What are the major fishery dependent data limitations/weaknesses and how could they be resolved? Define potential improvements and priorities for recommended improvements.
7. What recommendations do you have for prioritizing fishery-independent and fishery-dependent data collection improvements?
8. To what extent are fishery independent and fishery dependent data readily accessible to Center stock assessment scientists and to various external researchers who may wish to replicate NMFS stock assessments?
9. Identify the highest priority needs for improving fishery dependent and fishery independent data. Define potential improvements.

### Overarching Questions for Reviewers

- Relationship of current and planned fishery assessment data activities to Center fishery assessments mandates and requirements – is the Center doing the right things?
- Opportunities – are there opportunities that the Center should be pursuing in collecting and compiling fishery assessment data, including shared approaches with partners?
- Scientific/technical approach – are the Center's fishery data objectives adequate, and is the Center using the best suite of techniques and approaches to meet those objectives?
- Organization and priorities – is the Center's fishery data system properly organized to meet its mandates and is the allocation of resources among program appropriate?